

Université de Tartu
Collège des langues et des cultures étrangères
Département d'études romanes

Liina Maurer

LEXICOMÉTRIE POUR ANALYSER LES TWEETS D'EMMANUEL MACRON
Mémoire de licence

Sous la direction de
Marge Käsper

Tartu 2020

Table des matières

Introduction	4
1. Méthodes numériques, discours et politique	6
1.1. Méthodes numériques et discours	6
1.1.1. Humanités numériques	6
1.1.2. Analyses du discours françaises	7
1.1.3. Lexicométrie et présidents français	9
1.2. Politique et support numérique	11
1.2.1. Twitter et politique : <i>Twiplomacy</i>	11
1.2.2. Le format de Twitter	12
1.2.3. Les tweets politiques comme objet d'étude	13
2. Corpus d'analyse et premiers sondages	15
2.1. Constitution du corpus	15
2.1.1. Programmation pour obtenir le corpus	15
2.1.2. Nettoyage du corpus	16
2.2. Premiers sondages des données	16
2.2.1. Mots les plus fréquents	17
2.2.2. Les visuels	18
2.2.3. Évolutions dans les usages	19
3. Contextes pour discuter de l'ethos discursif de Macron	23
3.1. Notions et outils de travail	23
3.1.1. Modalité, cotexte et énonciation	23
3.1.2. AntConc comme outil d'analyse	25
3.2. Variations dans la dynamique de <i>devoir</i>	26
3.2.1. Dynamique sémantique entre 'changer' et 'continuer' ?	26
3.2.2. Une dynamique autre	29
3.3. Le « je » de Twitter	31

3.3.1. Usage le plus classique du dispositif.....	31
3.3.2. « Je veux » qui se définit.....	33
Conclusion	36
Références	40
Resümee	45
Annexe 1. Les résultats en chiffres et les calculs.....	47

Introduction

Dans ce mémoire, j'étudierai un corpus de tweets du président français Emmanuel Macron (@EmmanuelMacron). Comment le président utilise-t-il Twitter pour communiquer avec la société ? Comment formule-t-il ses messages ? Le but de ce mémoire est de discuter, dans une perspective socio-discursive, de Twitter comme mode de communication pour un président de la République en se concentrant sur la manière dont les mots qu'emploie Emmanuel Macron sur ce dispositif construisent et parlent de son ethos discursif de président.

Un défi spécifique relevé dans cette étude est d'y employer des outils lexicométriques, caractéristiques d'une partie de l'analyse française du discours¹, où ces outils ont été utilisés pour mettre en évidence des différences de style entre locuteurs. Tout en m'inspirant des études effectuées dans ce domaine, je vais interroger la manière de savoir comment décrire à l'aide des humanités numériques le profil discursif d'un président sans le comparer avec un autre.

Mon hypothèse est qu'à l'aide de programmes d'humanités numériques, il est possible de décrire un locuteur sans le comparer à quelqu'un d'autre, de définir des motifs répétitifs, de décrire les modèles de comportement discursif du locuteur et de modéliser les réactions ou expressions types probables à l'avenir.

J'ai choisi d'étudier les humanités numériques car il faut faire face au fait que la technologie prend le dessus, qu'il n'est plus nécessaire d'analyser à la main d'énormes corpus de texte, que les outils numériques peuvent simplifier le travail. Ce mémoire sera le premier dans le domaine pour les étudiants en sciences humaines dans le département d'études romanes, et de même, je ne connaissais pas les humanités numériques avant d'entreprendre ce travail. J'ai ainsi décidé d'y tester différents programmes et méthodes parce que je veux aider d'autres étudiants à l'avenir ou au moins attirer leur attention sur les nouvelles possibilités d'étudier les humanités de manière plus efficace. Pour cette raison, je consacre donc dans ce mémoire plus de temps à la présentation des outils lexicométriques que dans un mémoire ordinaire. L'analyse elle-même n'y sera qu'une voie ouverte pour les études à suivre.

¹ Désormais ADF.

Ce mémoire se présente en trois chapitres. Le premier chapitre est divisé en deux : dans la première partie je présenterai le cadre théorique des humanités numériques et de l'ADF, dans la deuxième partie je me rapprocherai de la politique et de Twitter, qui fournira le corpus de mes recherches.

Dans le deuxième chapitre, je me concentrerai sur l'obtention des données et leur traitement pour des analyses. Je présenterai en outre comment accéder à l'API Twitter et aux programmes lexicométriques qui peuvent aider d'autres étudiants à l'avenir. Le chapitre est conçu comme mon trajet dans ces outils, il serait donc facile à suivre et mes observations pourraient éventuellement aider les autres à prendre des décisions plus éclairées. Je discuterai également les premiers résultats fournis par les logiciels.

Le troisième chapitre se concentrera sur la précision et l'interprétation des données obtenues. J'approfondirai la dimension sémantique des récurrences trouvées et essayerai d'expliquer ce que ces résultats signifient en vue de caractériser l'ethos discursif de Macron sur Twitter. Quel est son « je » dans son énonciation ? J'espère que mes observations peuvent fournir des pistes pour interpréter les propos du président dans d'autres occurrences.

1. Méthodes numériques, discours et politique

Dans cette partie je présente ce que sont les humanités numériques et comment ce type de méthodologie a déjà été utilisé dans l'analyse française du discours et plus particulièrement en lexicométrie. À la fin du chapitre je présenterai ce qu'est la plateforme appelée Twitter où est collecté mon corpus, et ainsi que quelques études qui ont posé les tweets politiques comme objets de recherche.

1.1. Méthodes numériques et discours

Voici d'abord la présentation du cadrage théorique des humanités numériques et de l'ADF.

1.1.1. Humanités numériques

Avant de passer à mon sujet, il est nécessaire de clarifier quelle serait la bonne traduction en français pour ce qui se désigne en anglais par le terme *digital humanities*. Selon Marin Dacos et Pierre Mounier (2014) « Les humanités numériques recouvrent un ensemble de pratiques de recherche à l'intersection des technologies numériques et des différentes disciplines des sciences humaines ». Mais tout le monde n'est pas satisfait de l'utilisation du terme *les humanités numériques*.

L'Office québécois de la langue française (OQLF) privilégie l'utilisation de « sciences humaines numériques » et déconseille les expressions « humanités numériques » et « humanités digitales ». Pour l'OQLF, « humanités » est didactique et vieilli et « digitales » renvoie aux doigts. En guise de définition, le *Grand dictionnaire terminologique* propose « sciences humaines qui, à l'aide du potentiel des technologies de l'information, acquièrent une dimension nouvelle et deviennent accessibles de façon inédite ». (Perspectives SSF 2014)

Comme la note en effet un spécialiste français en *humanités numériques* Stéphane Pouyllau (2010) « [...] dans le monde des sciences humaines et sociales numériques (cette expression fait encore débat pour la version française du terme *digital humanities*) [...] ». The Humanities and Technology Camp en France et en Suisse, de 2010 à 2013, ont utilisé le terme *humanités numériques*, alors que d'autres francophones utilisent toujours *digital humanities* (Perspectives SSF 2014).

Laïd Bouzidi et Sabrina Boulesnane (2017), les auteurs d'une mise au point dans la revue *Les cahiers du numérique* intitulée « Les humanités numériques. L'évolution des usages et des pratiques » utilisent le terme *humanités numériques*.

J'ai aussi décidé d'utiliser *les humanités numériques* parce qu'en écrivant en français, c'est d'abord plus conséquent d'utiliser le terme français plutôt qu'un anglicisme, mais *sciences humaines numériques*, d'autre part, est trop long à utiliser.

Après avoir décidé quel terme utiliser, je vais expliquer ce que signifie les humanités numériques.

Selon Jean-Dominique Meunier (2017 : 20) les termes *humanities* en anglais et *humanités* en français ont des significations différentes. En anglais, *humanities* signifient généralement les sciences liées à la culture comme la sociologie, les langages, la littérature, etc. mais dans le monde francophone les humanités désignent généralement plutôt une tradition intellectuelle. En tout cas, c'est une approche pour systématiser des symboles comme des textes, des sons, des images qui sont traités par l'ordinateur. (Meunier 2017 : 20) « L'informatique appliquée aux sciences humaines voit avec le Web arriver un ensemble de pratiques et de problématiques nouvelles autour de la communication en ligne, et particulièrement avec le développement des médias sociaux » (Dacos & Mounier 2014 : 12). Les humanités numériques sont un mélange des sciences humaines et sociales qui se reconfigurent avec les outils de l'informatique, donc les humanités se réinventent par l'informatique et visent à changer la façon de faire de la recherche (Doduik 2019).

L'une des disciplines à la croisée de la société et de ses moyens de communication et intéressée par l'informatique depuis sa naissance, est l'analyse du discours.

1.1.2. Analyses du discours françaises

Dans les études sur le discours, même déjà depuis les années 1960, l'école française d'analyse du discours (ADF) s'intéresse aux diverses méthodes pour formaliser et modéliser la recherche. Alors que Michel Pêcheux (1969) a rêvé d'un outil automatique imaginaire pour détecter l'idéologie, les premiers travaux en lexicométrie politique ont été menés au Centre de la lexicométrie politique de St. Cloud (Maldidier 1969, Marcellesi 1971). Avec le temps, les méthodes basées sur la machine ont été discutées, développées et diversifiées (Guilhaumou 2002). Le principe de base de la

lexicométrie – de mesurer la fréquence des mots pour parler du social – a donné lieu à des analyses *textométriques* (Salem 1987), *logométriques* (Mayaffre 2004) ou encore *idéométriques* (Longhi *et al* 2017).

L'analyse du discours comme discipline, de fait a été établie en France en même temps et cependant de manière éparse par différents auteurs : (1) par le philosophe Michel Foucault qui a étudié d'une manière renouvelée l'histoire des idées, (2) par un théoricien de la société (et de l'idéologie surtout) Michel Pêcheux qui voulait intégrer aux sciences sociales la philosophie, l'histoire et la linguistique, et (3) par le linguiste Jean Dubois qui s'est concentré sur les usages sociaux du lexique (voir Käsper 2017 : 52-56).

Le lexicologue Jean Dubois a été en 1962 l'un des initiateurs des études lexicales discursives basées sur une conception du lexique articulé avec l'énoncé (les mots dans leurs usages). Il a essayé d'énumérer les usages lexicaux qui reflétaient les relations entre les différentes classes sociales pendant la Commune. Dans les années 70 et 80, Dubois a insisté sur l'idée que c'est la phrase qui devrait être considérée comme fondamentale pour l'analyse du lexique (et pas seulement les mots), parce que c'est dans l'usage (les mots insérés dans les phrases) que le mot acquiert la valeur d'énoncé. Ainsi, c'est au même moment que Denise Maltidier (1969) a étudié le vocabulaire de la guerre d'Algérie, et que Foucault (1969) a publié, la même année, « L'archéologie du savoir ».

Cependant, l'ADF a toujours mis l'accent sur la pensée critique pour discuter comment et dans quelle mesure ces analyses donnent accès au sens (voir à ce propos Guilhaumou (2002)). Ainsi, dans la discussion sur les humanités numériques, en montrant que pour la sémantique textuelle et l'analyse du discours « il y a un rôle à jouer et un dialogue interdisciplinaire nécessaire avec l'informatique ou les sciences de l'information et de la communication », Julien Longhi (2017) appelle « à ne pas oublier, dans la notion d'humanités numériques, le terme humanités, qui renvoie aux textes et documents ».

Comme l'explique l'un des théoriciens actuels centraux de l'ADF Dominique Maingueneau (2012), « l'intérêt de l'analyse du discours est d'appréhender le discours comme articulation de textes et de lieux sociaux. Son objet n'est ni l'organisation textuelle ni la situation de communication, mais ce qui les noue à travers un certain dispositif d'énonciation. » Cela veut dire en gros que dans l'étude de ce comment

s'exprime un locuteur, on s'intéresse à ses conditionnements et visées aussi bien sociologiques que textuelles. Comment comprendre et étudier plus précisément ces articulations, il y a des conceptions bien différentes en analyse du discours, l'analyse lexicométrique des vocabulaires en est une.

Dans ce mémoire, j'étudierai le vocabulaire du président en tant qu'institution qui se construit pour beaucoup à travers la parole (les mots énoncés) dans un médium et dispositif d'énonciation particulier, le Twitter. Mais avant d'aborder les spécificités de Twitter, considérons d'abord tradition des études sur la parole présidentielle.

1.1.3. Lexicométrie et présidents français

Les discours présidentiels ont toujours suscité des commentaires et par conséquent également des études approfondissant le sujet. De la même manière le président en tant qu'institution. Aujourd'hui, le travail le plus important en France et en lexicométrie a été produit par Damon Mayaffre (2004 ; 2007 ; 2018), qui a analysé les récurrences comparatives et les motifs de vocabulaire dans les discours publics de tous les présidents français depuis de Charles de Gaulle à Jacques Chirac (Mayaffre 2007), et plus tard jusqu'à Emmanuel Macron (Mayaffre 2020). Mayaffre (2004) fait remarquer par exemple les mots les plus fréquents de chaque président (*problème* pour Valéry Giscard d'Estaing, *civilisation* pour Georges Pompidou, *naturellement* pour Jacques Chirac, etc.) qui se démarquent comme sur-utilisés ou sous-utilisés lors de la comparaison avec les autres. Difficile certes de comparer Charles de Gaulle à Emmanuel Macron car les temps ont changé mais sur les statistiques de gros ensembles des données, ces études ont néanmoins constaté par exemple que lorsque Jacques Chirac, Nicolas Sarkozy et François Hollande ont essayé de réduire (par différents moyens langagiers) la distance avec l'interlocuteur, Macron joue avec les mots, comme Charles de Gaulle, et utilise délibérément des mots rares – « il refuse toute concession de langage qui ferait qu'il serait bien compris » (Gaboulaud & Lechevallier 2017).

Ce n'est pas bien sûr seulement les statistiques sur un certain nombre de mots sélectionnés, mais un grand nombre de données contrastées et leur analyse en groupes lexicaux très-élaborée au niveau sémantique qui donnent lieu à des interprétations en termes de profils socio-psychologiques et historiques des présidents étudiés dans cette

approche dite logométrique – sur le fond, c’est la logique de leurs arguments, leur *logos*, qui est étudiée par ces statistiques.

Or, l’argumentation se compose de différents aspects et l’une des conclusions de Mayaffre (2004 : 183-187) est qu’avec le temps, au détriment de *logos* les présidents ont commencé à travailler de plus en plus leur *ethos*. Dans ses recherches sur le discours politique, Longhi (2014 : 5-6) décrit cet effet comme constitution d’un *ethos* discursif et dans le projet #Idéo2017 (voir 1.2.3. ci-dessous) il semble se concentrer justement plutôt sur cet aspect-là.

Selon Marianne Doury (2016 : 120-121), il y a trois types de preuves rhétoriques sur lesquels le discours persuasif est basé : *ethos*, *pathos* et *logos*. *Ethos* discursif signifie le caractère ou l’image que l’orateur construit de lui-même dans son discours pour le spectateur ; *pathos* met l’accent sur les émotions et convictions du spectateur pour obtenir son soutien ; *logos* signifie logique, utilisation de faits et de preuves pour se rendre plus crédible (Doury 2016 : 120-142).

Le moyen le plus explicite pour repérer ce comment se construit l’*ethos* d’un locuteur est d’analyser son emploi des pronoms personnels. Les pronoms personnels sont par ailleurs un objet d’étude fréquent également dans le cadre général de l’ADF parce que le positionnement de l’énonciateur dans son discours est l’un des accents centraux dans les études de discours en français (Maingueneau 2016).

Quand on parle plus spécifiquement des statistiques étudiées quant à Macron, selon Mayaffre, Emmanuel Macron est le président qui utilise le plus de pronoms personnels de première personne comparé à tous les prétendants à l’année, dont principalement *je* et *nous* (Lorriaux 2017). Pour chaque 10 000 mot, il utilise dans sa campagne présidentielle 90 fois le *je*. C’est une stratégie où *nous* est extensible à l’infini et *je* est ce que lui a permis de se construire une stature, d’affirmer un *je* présidentiel (Lorriaux 2017).

Jean-Marc Leblanc (2016) s’est concentré sur les relations lexicales et syntaxiques des mots et des textes ; il a présenté le genre de discours présidentiels du Nouvel An. L’auteur insiste dans ce travail sur la nécessité de considérer le genre de discours en question, et les stratégies de communication liées à ce genre. La relation créée avec le public qui écoute étant toujours importante dans un discours du Nouvel An, il est possible de l’étudier par lexicométrie en se concentrant sur les différences dans

l'utilisation des pronoms personnels mentionnées par les présidents. Une démarche emblématique de la lexicométrie française illustrée dans Leblanc (2016 : 148) est d'ailleurs de faire représenter l'analyse sous forme d'analyse factorielle des correspondances qui indique les différences relatives d'utilisation des pronoms (je, moi / tu, toi / nous, etc.) entre les présidents.

L'usage des pronoms par Macron dans le corpus constitué dans le présent mémoire est donc l'un des aspects à considérer également ci-dessous.

1.2. Politique et support numérique

Il y a une tradition d'analyser les discours des présidents prononcés devant un public à l'écoute (Tulis 1987 – l'analyse rhétorique des discours des présidents américains) mais de nos jours, le support numérique de la parole est devenu de plus en plus important. Je prétends par conséquent que les médias sociaux sont aussi importants que les discours dans la transmission des images politiques.

1.2.1. Twitter et politique : *Twiplomacy*

Le développement technologique s'accélère. Dans le même rythme, nous avons changé notre façon de communiquer. Des milliards d'entre nous utilisent les réseaux sociaux chaque jour – l'une des plateformes est Twitter, qui est devenu une plate-forme également pour les effets de la « condensation sémantique » de l'information qui se produisent dans les tweets (voir Paveau 2012 ou Longhi 2013 : 25-32).

La diplomatie a évolué au fil du temps en raison des progrès technologiques (Chhabra 2020). Aujourd'hui, une grande majorité de personnalités politiques ou de partis politiques ont un compte Twitter pour signaler des informations, ce qui est en principe la fonction première de Twitter (Longhi 2014 : 1). En 2007, quand Barack Obama était sénateur, il est devenu le premier leader mondial à rejoindre Twitter ; en 2018, environ 97% des 193 États membres des Nations Unies avaient une forme de présence officielle sur Twitter (Wang 2019). Lorsqu'un leader mondial tweete, d'autres hommes politiques répondent au tweet ou le défendent, ou donnent simplement leur avis qui, à son tour, s'ajoute à la discussion en ligne qui façonne par conséquent l'opinion publique (Chhabra 2020).

Twitter a ainsi créé sa propre forme de diplomatie. *Twitter Diplomacy* ('diplomatie de Twitter') ou *Twiplomacy* est un mélange de diplomatie traditionnelle et numérique sur Twitter. Les médias sociaux franchissent les limites des traditions et sautent l'approche bureaucratique. (Chhabra 2020)

Selon Wang (2019), *Twiplomacy* offre trois avantages distincts : une accessibilité plus rapide à un public plus large, un contrôle des messages, ce qui signifie contourner les filtres des différentes agences, et des effets de réseau accélérés : en quelques minutes, les hommes politiques peuvent attirer l'attention et les réactions du public sur un tweet.

Twiplomacy encourage les citoyens à participer aux débats ou au processus décisionnel, la voix du public est clairement visible, ce qui aide à améliorer les relations et à développer la confiance. Aujourd'hui, de plus en plus de pays tentent de créer une image favorable dans les médias sociaux et donc dans l'esprit du peuple afin d'augmenter leur influence et leur prestige. Lorsqu'un homme politique communique avec le public, cela peut aider à créer une image d'homme politique de masse. (Chhabra 2020)

1.2.2. Le format de Twitter

Twitter est ce qui se passe dans le monde et ce dont les gens parlent en ce moment (About Twitter 2020).

Après une introduction sur la manière dont Twitter a changé ou est en train de changer la manière de faire de la politique, il est maintenant nécessaire de comprendre plus précisément ce qui est vraiment Twitter et comment il fonctionne.

Twitter est un service de microblogage en ligne permettant de diffuser des messages courts (280 caractères maximum)² entre groupes de destinataires. Le service a été conçu en 2006 par Evan Williams et Biz Stone. L'utilisateur tape un tweet via un mobile ou un ordinateur et l'envoie à une liste d'autres utilisateurs qui se sont inscrits pour recevoir les tweets de l'expéditeur. (Encyclopædia Britannica 2020)

Mais Twitter n'est pas seulement un service de courts messages. Il y a selon Marie-Anne Paveau (2012) tout un ensemble *des traits technodiscursifs du genre* qui

² La longueur du Tweet d'origine était de 140 caractères mais Twitter a évolué, la longueur maximale du Tweet est passée à 280 caractères (About Twitter 2020).

justifient que l'on décrive ces textes comme appartenant à un genre particulier, par exemple la contrainte des signes, les tweets en tweets³, transgressions grammaticales pour raccourcir le texte, des liens, des hashtags⁴, texte cliquable en différentes couleurs, aménagements typographiques comme sauter des espaces, etc.

Ce qui est également caractéristique de Twitter, c'est qu'un tweet est destiné à être retweeté, il initie la conversation et déclenche une série de réactions. Le fil du discours n'est plus présent, les hashtags importent des éléments contextuels et peuvent créer des interactions avec des tweets liés aux mêmes contextes. (Longhi 2014 : 3-4)

Il serait certes intéressant de suivre les tweets de Macron, suivre les fils du discours : qui et comment réagit, Macron réagit-il également ou à qui il répond. Cette méta-information est énorme et il y a d'autres logiciels spéciaux⁵ pour traiter ces informations. Dans ce mémoire, pour découvrir d'abord la voie de la lexicométrie, j'ai choisi de me concentrer uniquement sur les propos rédigés par Macron sur Twitter.

1.2.3. Les tweets politiques comme objet d'étude

Parmi les chercheurs français qui se sont penchés à étudier le fonctionnement des discours sur Twitter, Julien Longhi (2013) a relevé le tweet politique comme objet d'étude. Longhi (2013 : 1-2) considère le tweet politique comme un sous-genre de discours politique et ne le considère pas uniquement comme la « réduction de la pensée à 140 caractères ».

Tout en ayant commencé par une étude qualitative de tels tweets (Longhi 2013), lors des élections présidentielles françaises de 2017, Julien Longhi *et al* (2017) lancent une plate-forme #Idéo2017 pour analyser à l'aide d'outils de la lexicométrie ce que les candidats disent sur les réseaux sociaux. Le but dans ce projet était d'identifier les champs lexicaux et les thèmes dominants des différents candidats et ainsi d'établir « le profil linguistique » des idées des candidats et de ce qui les différencie.

Dans cet ensemble, #Idéo2017 ne donne pas une longue analyse de ce qui est Macron, sauf pour les visuels, parce que l'objectif était d'offrir à tout le monde un accès gratuit

³ Quand une personne fait référence à un autre tweet.

⁴ Hashtag est un mot ou une phrase de mots écrits sans espaces commençant par un signe d'hashtag (#) pour identifier les messages sur un sujet spécifique.

⁵ L'un de ces outils est Gephi qui a également été utilisé pour observer les candidats à la présidentielle sur Twitter en 2017 (Gaumont *et al* 2018).

aux tweets et un moyen facile de les filtrer en fonction du sujet d'intérêt – chacun peut faire ses propres recherches et analyses. (#Idéo2017 plateforme 2017) Le site propose quatre types d'analyses à travers 13 mots sélectionnés en français par exemple *France*, *état*, *sécurité*, etc. La première analyse propose d'identifier l'utilisation du mot choisi par les différents candidats. Les résultats sont présentés sous forme de deux graphiques : l'un pour le calcul des spécificités (sous-utilisation et sur-utilisation du mot par les candidats), et l'autre pour la fréquence d'utilisation du mot par les candidats. (#Idéo2017 plateforme 2017)

Alors que Julien Longhi a étudié les tweets des candidats pendant les élections présidentielles, dans ce mémoire, j'étudierai un corpus des tweets de Macron comme un seul sujet et la période de l'étude commence lorsque Macron est déjà devenu président de la France.

2. Corpus d'analyse et premiers sondages

Le deuxième chapitre parle plus spécifiquement de mon corpus : comment j'ai collecté, filtré et analysé les tweets d'Emmanuel Macron. Je présenterai également les premiers logiciels lexicométriques que j'ai utilisés pour sonder le corpus et avoir les premiers résultats à interpréter.

2.1. Constitution du corpus

Pour me doter d'un corpus à étudier, il était nécessaire d'obtenir d'abord les tweets de Twitter. Pour accéder Twitter API, il est nécessaire de demander un compte développeur Twitter. Depuis 2019, Twitter a par ailleurs imposé une limite : seuls les 3200 tweets peuvent être téléchargés à la fois. (Get Tweet Timelines 2020)

2.1.1. Programmation pour obtenir le corpus

Je vais donner une instruction précise sur ce que j'ai fait pour obtenir les tweets de Twitter. J'ai suivi les instructions sur la page développeur Twitter « Get Tweet timelines (2020) ». Il existe plusieurs façons d'obtenir les données mais je vais décrire mon expérience. J'ai utilisé R ('The R Project for Statistical Computing') qui est un environnement logiciel libre pour le calcul statistique et les graphiques (R Project 2020).

1. Tout d'abord, c'était nécessaire d'installer les packages intégrés de R « rtweet » et « writexl » pour générer un fichier xls – cela signifie que j'ai pu obtenir les tweets sous forme de fichier Excel Microsoft.
2. Après, j'ai utilisé la commande « install.package ».
3. J'ai défini la bibliothèque de R avec la commande « bibliothèque ».
4. J'ai utilisé la fonction « get_timelines » avec deux paramètres : le nom d'utilisateur Twitter du compte d'Emmanuel Macron et le nombre de tweets possible (3200).
5. Alors j'ai utilisé la fonction « write_xlsx » avec trois paramètres : (1) j'ai enregistré mes résultats précédents, (2) où je souhaitais enregistrer mes résultats finaux, et enfin (3) j'ai écrit « true » ('vrai en anglais') parce que je voulais avoir les noms des colonnes dans le fichier Microsoft Excel.

Voici dans la figure 1 le code rédigé pour obtenir les tweets d'Emmanuel Macron.

```
install.packages("rtweet")
library(rtweet)
timeline <- get_timelines("emmanuelmacron",n=3200)
|
install.packages("writexl")
library(writexl)
write_xlsx(x = timeline, path = "c:/Users/ L.M. /Desktop/macron.xls", col_names = TRUE)
```

Figure 1. Le code Twitter get_timelines

2.1.2. Nettoyage du corpus

J'ai reçu 3200 derniers tweets (tiré en novembre 2019) d'Emmanuel Macron. Pour me concentrer principalement sur Emmanuel Macron, il fallait sélectionner le texte qu'il a probablement écrit par lui-même.

Parce que c'est le compte public du président, une certaine quantité de texte est rédigé ou éditée par son équipe de relations publiques. Mais en se concentrant sur le tweet politique comme Longhi (2014 : 3-10), ce n'est pas important de savoir qui a écrit le tweet, ce qui compte, c'est l'éthos discursif qu'il crée. Mayaffre (2004) a étudié des discours présidentiels qui ont tout aussi bien pu être rédigés par des conseillers, mais à partir de ces discours les gens ont formé leur opinion sur le président.

Les données sont fournies sous forme de tableau Microsoft Excel où chaque ligne est un tweet et chaque colonne contient des informations sur ce tweet. Par exemple la colonne X est nommée « retweet » où dans chaque ligne est écrit « oui » ou « non », donc il est possible de filtrer un texte que le président n'a pas écrit lui-même. J'ai également supprimé les tweets qui ont été rédigés dans une autre langue car ces tweets ont aussi été traduits en français.

Après filtrage, il m'est resté 2306 tweets entre le 12 décembre 2017 et le 11 décembre 2019 pour analyser un total de 73594 jetons de mot⁶.

2.2. Premiers sondages des données

Ce sous-chapitre est conçu comme mon trajet, qui continue dans le chapitre suivant, comment j'ai utilisé trois logiciels différents des humanités numériques. Je vais

⁶ Le nombre de jeton de mot signifie le nombre total de mots dans un texte.

présenter brièvement les programmes, ce que j'en ai fait et mon opinion en tant qu'utilisateur.

2.2.1. Mots les plus fréquents

Initialement, c'est avant tout Lexico qui était destiné à être utilisé car, élaboré au départ à l'École normale supérieure de Fontenay-Saint-Cloud par André Salem (Patin & Megale 2018 : 124), c'est l'un des logiciels emblématiques dans la lexicométrie française. J'ai décidé d'utiliser Lexico 5 qui est la dernière et devrait être alors la version de ce logiciel la plus à jour.

Lexico est un logiciel pour les méthodes d'analyse textométrique. Il fonctionne sur Windows et émulateur Windows sous Mac. Lexico 5 contient six outils : des outils de segmentation, de concordances, de partitions, de segments répétés, d'analyse des spécificités et d'analyse des correspondances et il comprend 7 langues. (Lexico 5 2019)

J'ai téléchargé le fichier txt, avec tous les tweets d'Emmanuel Macron, dans Lexico et j'ai essayé de tester différents outils. Tout d'abord, j'ai commencé à regarder la fréquence des mots. Dans les discours présidentiels étudiés par Mayaffre (2004 : 25), depuis de Gaulle jusqu'à Chirac, *France* était le mot plein⁷ le plus utilisé et c'était aussi le mot le plus utilisé par Macron dans mon corpus : il l'a utilisé 370 fois, ce qui est donc bien logique d'un Président de la France. Le deuxième mot le plus utilisé par Giscard était *problème* alors que pour tous les autres présidents c'était *pays* – ces variations sont une façon de décrire la caractéristique (Mayaffre 2004 : 25).

J'ai commencé à tester l'outil de concordances⁸ dans Lexico et on peut voir les premiers résultats sur la figure 2. Des dix mots pleins de Macron selon Lexico 5 par ordre décroissant étaient *France*, *pays*, *français*, *monde*, *ensemble*, *ans*, *aujourd'hui*, *république*, *européenne* et *nouvelle*. Sur la gauche de la figure, on peut voir les mots les plus fréquemment utilisés de Macron sans utiliser des mots d'arrêts⁹. Dans la boîte

⁷ Les mots pleins selon Mayaffre (2004 : 25) étaient les noms et les adjectifs.

⁸ L'outil de concordance détecte le mot ou les segments de recherche dans le texte et montre ce qui est avant et après l'élément recherché.

⁹ Le mot d'arrêt ici signifie que le logiciel laisse quelques mots sans compter, par exemple les conjonctions.

principale se trouvent les concordances de *France* et des mots qui sont utilisés précédemment et après.

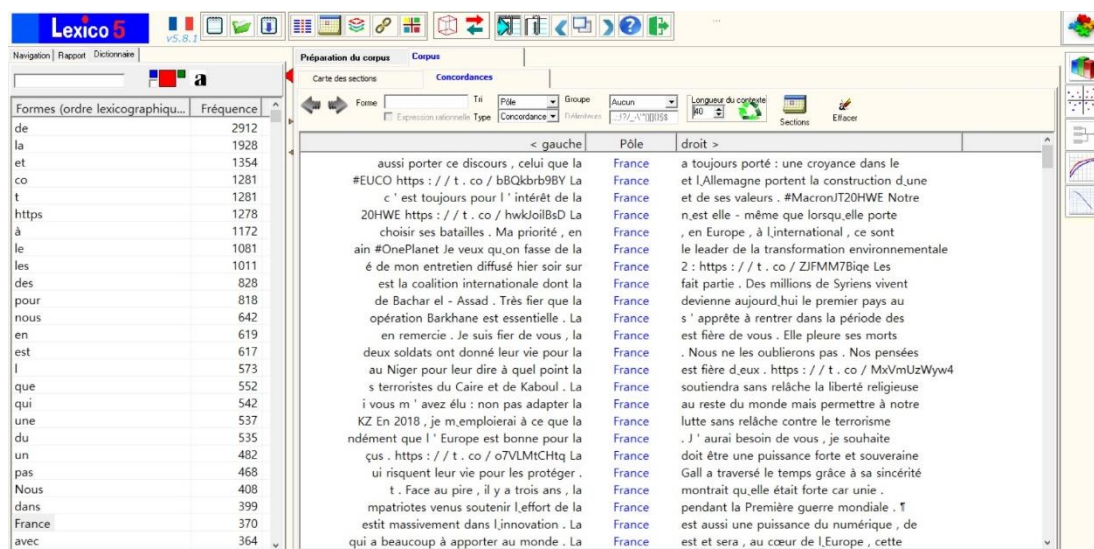


Figure 2. Les mots les plus fréquents dans le corpus et les concordances de *France* avec Lexico 5

Mayaffre (2004 : 38) a également présenté les verbes les plus utilisés : *avoir*, *être* et *faire* étaient prioritaires pour presque tous les présidents, et selon Lexico, Macron n'était pas une exception.

Pour une analyse plus approfondie, Lexico 5 était toutefois difficile à utiliser parce qu'il manque d'informations, il n'y a pas de manuel non plus. Le logiciel lui-même a été publié mais quand je l'ai utilisé au printemps 2020, la page d'accueil était encore en construction. Lexico est à usage professionnel mais sans manuel, en tant que débutante, je n'ai pas pu profiter de tout ce qu'il offre, par exemple des visuels. Mais les visuels sont juste ce qu'il faut pour comprendre. Les visuels aident à trouver des informations que les tableaux de nombres ne révèlent pas nécessairement.

Il existe de nombreux logiciels similaires, mais l'un des plus faciles à utiliser en tant qu'étudiante, et que je recommande aux autres, est Voyant Tools.

2.2.2. Les visuels

Voyant Tools est un environnement de lecture et d'analyse basé sur le Web pour les textes numériques de Stéfan Sinclair et Geoffrey Rockwell (2016). Il est conçu pour être vraiment facile avec son propre texte dans une variété de formats, y compris du

texte brut, HTML, XML, PDF, RTF et MS Word. Il contient 5 outils principaux : cirrus, lecteur, les tendances, sommaire, contextes et de nombreux outils supplémentaires. Il est disponible en 10 langues. (Voyant Tools Help 2016)

Parmi les programmes d'analyse de texte que j'ai testés, Voyant Tools était le plus facile à utiliser. C'est une très bonne plateforme pour commencer en tant que débutante, parce qu'on n'a pas besoin de la télécharger et elle crée automatiquement de nombreuses figures esthétiquement belles, par exemple un nuage avec les mots les plus fréquemment utilisés. Pour donner un exemple, on peut voir la figure 3. Plus le mot est gros, plus le mot est utilisé.



Figure 3. 25 mots les plus fréquemment utilisés de Macron visualisé par Voyant Tools

Le problème avec Voyant Tools est cependant que rien n'est modifiable. Les visuels sont créés automatiquement, on peut choisir la quantité de données mais l'apparence du visuel n'est pas modifiable.

2.2.3. Évolutions dans les usages

Après avoir reçu la liste des mots les plus fréquents, j'ai commencé à chercher des mots ou des segments sur lesquels il serait plus intéressant de se concentrer. Les mots les plus fréquents comme *France*, *Europe*, *monde*, *avoir*, *être*, etc. sont dans chaque discours présidentiel ainsi que dans de nombreux tweets. Je voulais trouver quelque chose de caractéristique pour Macron qui le différencie des autres présidents, aussi quelque chose de plus inattendu mais suffisamment utilisé pour avoir assez de données.

Ayant regardé la liste des mots les plus fréquemment utilisés dans Lexico, j'ai remarqué que le quatrième verbe le plus utilisé par Macron est devoir, dont dans mon corpus surtout la forme plurielle de la première personne *devons*. Selon Mayaffre (2004 : 38), aucun président ne l'a utilisée aussi souvent. Donc, j'ai décidé de tester *devons* dans les tweets de Macron.

Mayaffre (2004 : 38-39) a découvert que le verbe *dire* monte et descend au fil du temps, j'ai donc voulu voir les changements de la forme *devons*. Voyant Tools est efficace pour créer des visuels indicatifs car il segmente automatiquement le document en parties égales ce qui rend le corpus comparable. Pour donner un exemple, on peut voir la figure 4 où Voyant Tools montre les changements de fréquence du mot recherché au fil du temps.

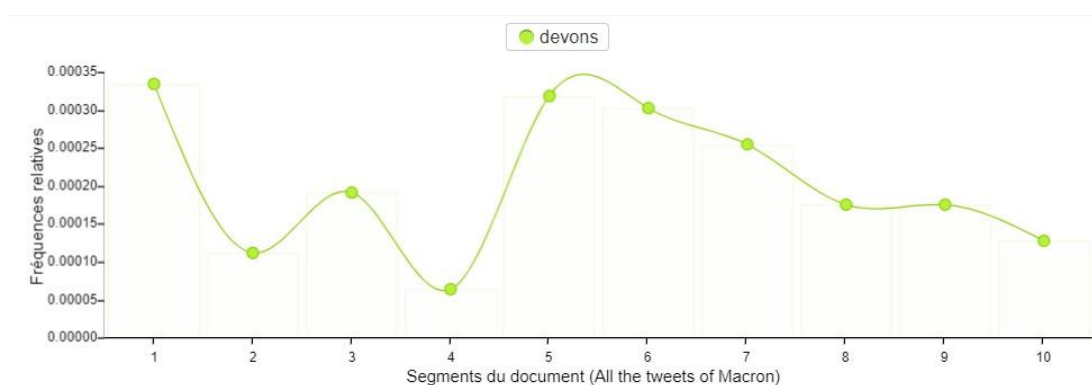


Figure 4. Fréquences relatives de *devons* avec Voyant Tools

Ce que ce visuel a donné, c'est l'assurance que la fréquence de *devons* a changé au fil du temps. Il donne les directions des parties les plus actives, par exemple, on peut voir qu'entre le segment 4 et 5 il y a une grande augmentation. Je commençais à me demander pourquoi. Pour comprendre la cause de la montée et de la chute, j'ai commencé à regarder ce qui s'est passé pendant la période couverte par mon corpus.

L'un des grands événements a été le mouvement des Gilets jaunes (Lehut 2019) qui au début se sont battus contre la hausse de la taxe sur les carburants mais après contre Emmanuel Macron et sa politique ; leur première marche a été le 17 novembre 2018 – au milieu de ma période de recherche. J'ai décidé de diviser mon corpus en deux : avant et depuis le mouvement (parce que ce n'est pas fini) pour comparer si Macron a changé son style de tweet depuis le début du mouvement.

J'ai décidé rassembler le calendrier exact du corpus, le nombre de tweets et mon mot recherché *devons* pour avoir une vision claire de ce qui s'est passé et quand. On peut voir la figure 5.

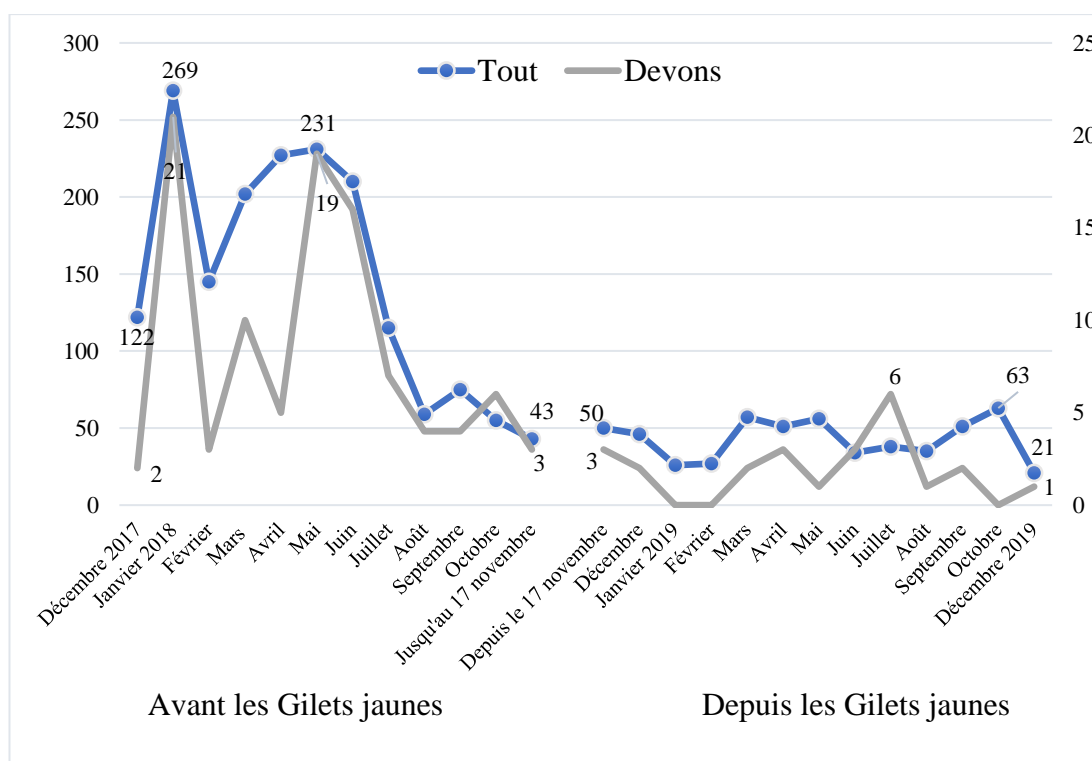


Figure 5. Le nombre de tweets et la fréquence du mot *devons*

Je pourrais écrire les résultats mathématiquement, ce qui les rendrait plus difficile à lire, mais ce qui est important, c'est que ces résultats sont très intéressants. Le période avant le mouvement a duré 340 jours et pendant ce temps Macron a tweeté 1751 fois, en total de 50027 mots. À partir du 17 novembre 2018, date de la première marche des Gilets jaunes, la deuxième période a commencé et cela a duré 390 jours, pendant ce temps, Macron a tweeté 555 fois, en total de 23467 mots. Si on divise le nombre de tweets par le nombre de jours, on obtient le résultat en moyenne de 5,15 tweets par jour pendant la première période et 1,423 pendant la deuxième période. Si on divise 5,15 par 1,423, on obtient le résultat que depuis le mouvement des Gilets jaunes, Macron a commencé à tweeter 3,619 fois moins.

Ce qui peut également être vu sur la figure, c'est que la fréquence des tweets baisse significativement au cours de l'été 2018 et il en va de même de l'utilisation du mot *devons*. Les problèmes avaient commencé avant la première marche – c'est pourquoi la baisse a commencé avant le 17 novembre. La popularité du président a diminué,

alors il est devenu plus prudent sur ce qu'il a dit. Compte tenu de la situation politique, on comprend pourquoi le nombre de tweets a diminué. Macron a probablement réévalué ses actions, il a pensé plus à quoi dire ou à ne pas dire pour ne pas aggraver la situation.

Pour comprendre plus précisément comment il a changé, on doit regarder le contenu des tweets. Un article d'Olivier Bénis et Cyril Graziani (2020) reste en dehors de la période de mon corpus mais décrit comment le coronavirus a changé la façon de tweeter de Macron : il répond à d'autres tweets, ce qu'il ne faisait presque jamais et il est un peu moins formel, hésite moins à montrer une proximité. Cela montre comment Macron agit pendant la crise et prouve que les changements dans l'environnement changent le style de tweet de Macron.

Les outils de Voyant Tools m'ont donc donné un premier aperçu de ce qui s'est passé mais pour aller plus en détails dans mon analyse des tweets, il faudrait quand même un outil lexicométrique plus au point. Comme mon expérience avec Lexico s'était avérée pas très productif, au détriment de Lexico, j'ai décidé d'utiliser un autre logiciel classique, certes pas français, mais comparable dans ses fonctionnalités de base à Lexico – AntConc.

3. Contextes pour discuter de l’ethos discursif de Macron

Le troisième chapitre se concentre sur l'analyse plus approfondie des tweets et sur leur interprétation, afin de caractériser, sur la base des résultats de ces analyses lexicométriques, ce qu'on pourrait appeler le style d’Emmanuel Macron sur Twitter – son ethos discursif dans ce contexte. Est-ce que son style change au cours du corpus, et si, comment est-ce qu’il change sur le fond ? Je commence par présenter mes outils théoriques et pratiques, pour mener ensuite dans un premier lieu une analyse plus précise des usages du mot clé relevé dans les sondages – *devons* – et puis une analyse qui se consacre plus précisément sur le « je » d’Emmanuel Macron sur Twitter.

3.1. Notions et outils de travail

Voici d’abord la présentation du cadrage théorique et du logiciel utilisé pour les analyses de ce chapitre.

3.1.1. Modalité, cotexte et énonciation

Si les logiciels ont relevé comme l’un des mots les plus récurrents dans le corpus une forme du verbe *devoir*, il fallait se demander quel est plus particulièrement l’essence et la portée de ce mot.

Quant à la nature du verbe, le verbe *devoir* est un verbe modal, dont le rôle est en principe de nuancer la portée du verbe principal dans la phrase.

Selon Nicole Le Querler (2004 : 646, 650 in Narusk 2014 : 5-6), « la modalité est l’expression de l’attitude du locuteur par rapport au contenu propositionnel de son énoncé » et il y a plusieurs façons de l’exprimer : avec les adverbes, les temps verbaux, les subordonnées et, notamment, les verbes modaux, dont le verbe *devoir* fait partie.

Dans une étude qui se concentre dans une approche discursive sur le verbe *devoir* dans les documents politiques portant sur le changement climatique, Engelsén (2014 : 30-34) résume selon Hans Kronning (1996) et Le Querler (1996) trois types de modalités pour le verbe *devoir* : (1) modalité épistémique, qui se réfère au jugement de ce que l’énonciateur sait, croit ou suppose ; (2) modalité déontique, qui indique l’obligation, volonté ou souhait exprimé par l’énonciateur dans un rapport intersubjectif ; et (3)

modalité aléthique, ce qui signifie la relation entre le contenu propositionnel et la vérité nécessaire.

Je vais essayer de tenir compte de ces distinctions en analysant les propos de Macron, mais les tweets sont courts et surtout dans la première forme plurielle *nous*, il n'est pas toujours évident de distinguer si le *nous devons* implique une obligation ou souhait adressé au public ou bien il s'agit plutôt d'un jugement exprimé par Macron dans une « stratégie où *nous* est extensible à l'infini » (Lorriaux 2017).

C'est pourquoi j'ai décidé de m'appuyer sur ce que les logiciels me fournissent dans le cotexte immédiat – les verbes qui suivent la forme *devons* – et d'analyser la sémantique des usages de la forme *devons* de Macron par cette voie (tout en tenant compte du cotexte élargi du tweet aussi). Si dans les matériaux de Engelsen (2014 : 85), 81 occurrences sur 96 se situaient dans la modalité déontique et pour préciser ses usages du verbe *devoir*, l'auteur a élaboré une approche sémantique spécifique pour son contexte, je vais proposer comme axe spécifique du contexte de Macron la question d'une dynamique entre changements et continuité. Normalement, quand on parle de la modalité et des verbes modaux, au sens strict, en sémantique il est question avant tout de la possibilité ou de la nécessité exprimée dans l'énoncé (voir Vetters 2012 : 31-47), mais dans le discours de Macron, c'est la question de changements à entamer ou de situation à maintenir qui semblait mieux résumer les thématiques dont il parlait en employant le mot *devons*.

D'autre part, pour revenir à élucider la stratégie évoquée par Mayaffre quant à Macron dans Lorriaux (2017), où *nous* serait extensible à l'infini et *je* serait ce que lui aurait permis de se construire une stature, « d'affirmer un *je* présidentiel », j'ai toutefois voulu étudier la construction de cette « stature ». Si le *nous* de Macron ne semblait pas facilement analysable, comme il a été dit dans la partie introductive, le pronom *je* est le moyen le plus explicite – et souvent exploité en analyse du discours à cette fin – pour analyser le rapport qu'a le locuteur à son énoncé. Alors la deuxième partie de ce chapitre examinera les segments de *je* de Macron précisés en cotextes. Pour délimiter les possibles occurrences à analyser, je vais me concentrer sur un segment qui a été indiqué comme fréquent chez Macron (Lorriaux 2017) : *je veux*. Par ailleurs, *vouloir* est aussi un verbe modal, et, même s'il n'est pas central dans les typologies de la

modalité à côté de *pouvoir* et *devoir*, ce verbe est très intéressant en ce comment il donne accès aux idées exprimées par Macron.

En dehors de la sémantique, il faut préciser encore que c'est aussi la modalité au sens le plus large comme *manière de dire* qui intéresse l'analyse du discours, pour tenir compte de la variabilité des dits dues aux dispositifs d'énonciation. C'est pourquoi, même si je vais me concentrer sur le lexique, je vais discuter aussi d'autres types de moyens – les statistiques précisées, les visuels et autres moyens employés – pour étudier ce rapport observable de Macron dans son énonciation sur Twitter.

3.1.2. AntConc comme outil d'analyse

AntConc est un outil multiplateforme de Anthony Laurence pour la réalisation de recherches linguistiques de corpus et d'apprentissage basé sur les données, fait pour fonctionner sur n'importe quel ordinateur exécutant Windows, MacOS X et Linux (Manuel AntConc 2019 : 1). Il contient sept outils : outil de concordance, de tracé de concordance, d'affichage des fichiers, des clusters, des colocalisations, une liste des mots et des mots clés (Manuel AntConc 2019 : 2). AntConc est entièrement compatible unicode : il comprend toutes les langues européennes et asiatiques. Il offre également de nombreuses options, par exemple pour modifier les polices ou les couleurs du texte ou les résultats qu'on souhaite afficher. (Manuel AntConc 2019 : 8)

Les outils que j'ai surtout utilisés pour mes recherches étaient concordance et clusters¹⁰.

La modalité se concentre sur le contexte mais pour utiliser les possibilités des humanités numériques, j'ai donc décidé de me concentrer sur le cotexte. Avec l'aide de segments répétés et de concordances, j'essayerai de trouver des modèles de *devons* – quels mots vient après pour voir ce que Macron suggère ou veut faire faire, changer, conserver, garder, etc. On peut voir un exemple dans la figure 6. Dans la boîte principale, on voit ce qui vient après le mot recherché et à gauche il y a le rang de la fréquence et combien de fois ce segment est-il utilisé.

¹⁰ *Clusters* dans AntConc est le même que les segments répétés dans Lexico et à partir de maintenant j'utiliserai le terme français. Le but de l'outil est de rechercher des segments répétés de mots et de trouver des modèles dans la langue.

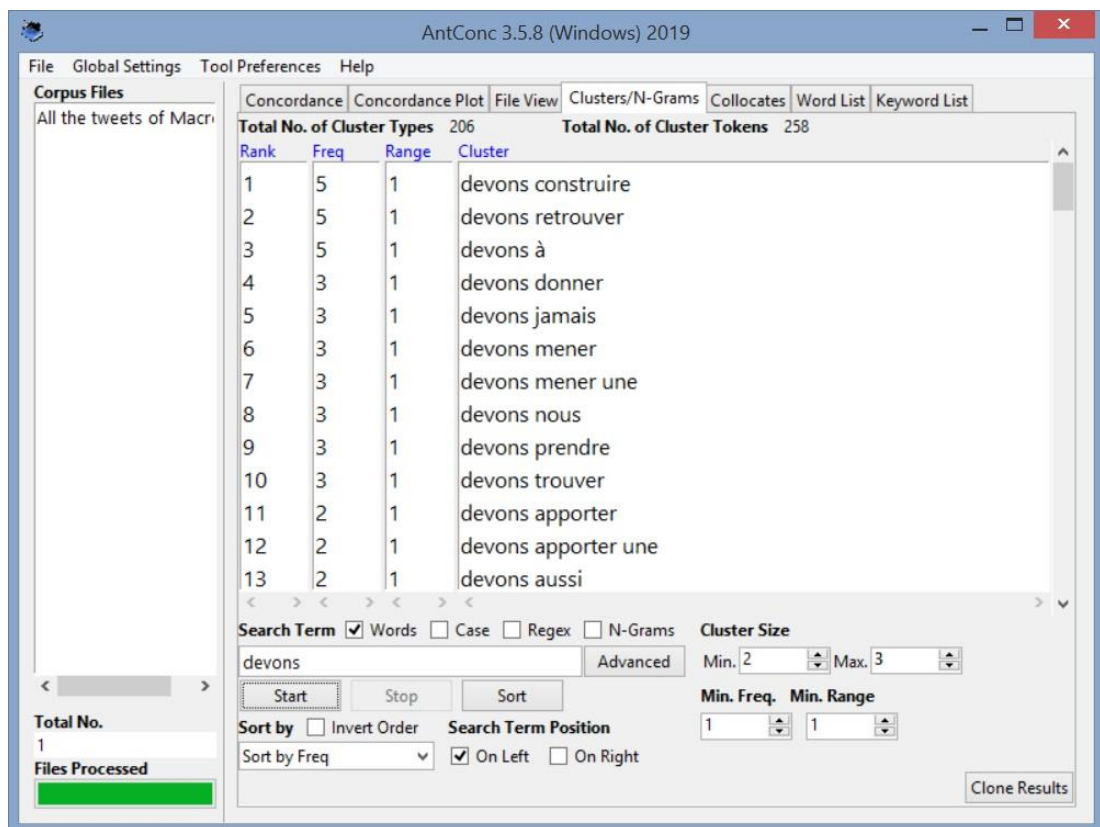


Figure 6. Segments répétés de *devons* avec AntConc

Une façon d'analyser est de donc regrouper et AntConc l'a fait pour moi. J'ai voulu savoir si Macron a changé en ce qu'il entend par ses évocations de devoirs, ou plus précisément par ses évocations de *devons*. Pour cela, j'ai comparé les segments avant et depuis le mouvement des Gilets jaunes pour voir les changements.

3.2. Variations dans la dynamique de *devoir*

Voici d'abord la présentation d'une analyse plus précise en utilisant un axe sémantique.

3.2.1. Dynamique sémantique entre 'changer' et 'continuer' ?

En lisant les concordances de ces groupes j'ai tenté de les diviser à mon tour en deux catégories : *nous devons changer* – 'changer' – ce qui signifie que quelque chose doit changer, être différent de la situation actuelle, doit agir, etc. ; et *nous devons continuer* – 'continuer' – ce qui signifie qu'il faut continuer, garder, prendre en compte les valeurs, etc. (pour les résultats exacts de l'analyse effectuée, voir l'annexe 1).

Dans certains cas, la distribution des exemples entre 'changer' et 'continuer' se faisait assez facilement, voyons le premier exemple :

Ex. 1. Nous devons renouer avec le goût du risque et renoncer à la peur de l'échec. Nous ne sommes pas une Nation d'assis, faites par des femmes et des hommes qui n'ont pas osé. Au contraire. <https://t.co/zLQgNgN9Je>. (Macron 30/01/2018)

C'est un 'devons changer' pour moi parce que le tweet souligne la nécessité de changer notre façon de penser et *oser risquer*, en ce moment, *nous* ne sommes pas qui nous sommes vraiment.

Ex. 2. Merci d'avoir refusé le fatalisme. Nous devons tenir nos engagements sur le climat. Ensemble, nous allons gagner la bataille. #OnePlanet #OnePlanetSummit <https://t.co/UHw0xGa54a>. (Macron 12/12/2017)

Cet exemple est 'devons continuer' parce qu'il met l'accent sur les actions à poursuivre pour garder les valeurs de prendre soin de l'environnement, de la responsabilité et de penser à l'avenir. « Gagner la bataille » et « refusé le fatalisme » (c'est-à-dire ne pas se contenter de ce que la fatalité apporte) peuvent certes être aussi interprétés comme 'devons changer' dans l'ensemble de l'énoncé, mais puisque le mot *devons* se trouve associé aux engagements déjà pris, dans cet exemple j'interprète la nécessité indiquée par Macron comme 'devons continuer'.

Quand on se concentre sur les polarités dans le discours de Macron, il faut noter en fait que Macron a un tic de langage « en même temps », qu'il connaît, dont il se plaisante même, mais qu'il utilise toujours « car ça signifie que l'on prend en compte des principes qui paraissaient opposés » (Mongaillard 2018). Cette expression elle-même n'apparaît que 9 fois dans mon corpus mais ce tic indique sur le fond que dans l'ensemble il équilibre toujours des deux côtés. Il pense toujours d'une part et d'autre part, ce qui fait que même si ses formulations semblaient changer, le tableau final de mes distributions quant à ses incitations à 'changer' ou 'continuer' par le mot *devons* n'a pas montré de grands changements ce qui change réellement.

Déjà, certains tweets étaient difficiles à interpréter sur cet axe de changements.

Quant au sens plus précis de *devoir*, la fonction de *devons* n'est pas toujours de donner des ordres, ni même de modaliser un verbe. Dans un sens qui est considéré complètement à part par Engelsén (2014) car non-modal, le verbe *devoir* combiné à la préposition *à* peut aussi signifier gratitude ou humilité. Par exemple Macron a rendu hommage à ceux qui se battent pour la paix. Il a utilisé *devons* dans ce sens plusieurs

fois, mais il mélange aussi souvent l'humilité avec une suggestion cachée ou un souhait. On peut regarder l'exemple 3.

Ex. 3. L'Europe doit être en mesure de se défendre et de protéger ses États membres. C'est ce que nous sommes en train de faire. C'est ce que nous nous devons les uns les autres. <https://t.co/8Cyxrpf7H>. (Macron 30/08/2018)

Je le comprends comme son souhait de ce comment l'Europe devrait être et je n'ai donc pas écarté ces exemples de mon analyse. Il a souligné le besoin de protection en disant que nous nous devons les uns aux autres. J'ai interprété la gratitude ou l'humilité comme 'devons continuer' car c'est quelque chose qu'il ne faut pas oublier, il faut continuer à faire, merci ne peut pas finir. Mais c'est une situation complexe parce que tout le sens de la phrase pourrait être qu'il faut changer quelque chose. Peut-être il pensait que l'Europe devait faire mieux pour se protéger ou que les États membres devraient augmenter le budget des dépenses de défense – le contexte précis n'est pas toujours visible.

La question de qui doit m'a fait réfléchir aussi à cotexte à gauche. Pour cette raison, j'ai aussi décidé de tester en parallèle à *devons* la forme *doit* – la troisième forme singulière de *devoir* car avant *devons* il y a généralement *nous* mais devant *doit* on peut avoir *on* mais aussi beaucoup d'autres mots – je voulais savoir qui doit et quoi – qui est nous.

Quand on regarde qui est ce qui *doit* dans les tweets de Macron, il mentionne l'Europe, la France, l'Iran, la République, chacun, on, la Réunion, le français, la francophonie, la relation entre quelqu'un, le budget, l'égalité, la méthode, la réponse, notre avenir, etc. Mais il n'a pas commencé à mentionner quelqu'un ou quelque chose plus ou moins depuis le mouvement. Il y avait trop d'options différentes de qui pour un trop petit nombre de *doit* – il n'y a pas assez de données pour regrouper. On peut conclure que Macron mentionne divers sujets et donne l'impression qu'il communique avec tout le monde.

Mes résultats peuvent sembler modestes, mais compte tenu des commentaires sur son tic de langage *en même temps* et des distributions de *devons* finalement presque équilibré entre changement et continuité. Cela peut en fait être son style qui se fait exprès de ne pas répéter la même chose pour ne pas donner la possibilité de remarquer quelque chose de particulier. Il est compréhensible qu'on puisse comprendre le

contenu différemment, donc je ne prétends pas que tout le monde obtienne exactement les mêmes conclusions. La distribution sur l'axe sémantique entre changer et continuer semblait une bonne idée, mais en réalité Macron est très difficile à interpréter, ses messages ont souvent plusieurs idées à la fois, *en même temps* fait partie de son style et de sa stratégie.

On pourrait se demander pourquoi j'ai choisi ces mots et segments et pas d'autres. La raison est d'une signification statistique. Parce que mon corpus est petit, la fréquence des mots est petite, il y a donc très peu de segments récurrents. J'ai testé différents mots mais si la fréquence est inférieure à 0,000... cela ne suffit pas pour faire une hypothèse. Mais ce que j'ai étudié, je peux le confirmer avec des calculs.

3.2.2. Une dynamique autre

En regardant mes résultats, on pouvait finalement voir un équilibre entre 'changer' et 'continuer'. Il n'y en a pas avant et l'autre depuis le mouvement – les deux sont représentés et il semble que presque à part égale dans les deux périodes étudiées. Est-ce qu'il n'y avait pas de changement du tout dans le style, sauf celui de la fréquence de tweeter que j'ai montré dans la partie 2.2.3 ?

Tout d'abord, j'ai voulu vérifier la longueur des tweets. Dans le chapitre 2.2.3 j'ai présenté le nombre de tweets et le calendrier des deux périodes, mais pour faciliter la lecture, je présenterai à nouveau tous les chiffres et les résultats des calculs avec le tableau numéro 1.

Tableau 1. La fréquence et durée des tweets

		A	B
-	Plage de temps	Avant le mouvement des Gilets jaunes	Depuis le mouvement des Gilets jaunes
1	Nombre de jours de la période	340	390
2	Nombre de tweets	1751	555
3	Nombre de jetons de mot	50027	23467
-	Nombre de tweets par jour	$A2/A1=5,15$	$B2/B1=1,423$
-	Jetons de mot par tweet	$A3/A2=28,571$	$B3/B2=42,283$

Il est vrai qu'au début, il a plus tweeté, donc il a utilisé plus de mots, on a l'impression que pendant la deuxième période Macron est resté silencieux. Mais si l'on divise le

nombre de jetons de mot par le nombre de tweets on peut voir que Macron a commencé à écrire davantage dans chaque tweet. Les calculs confirment que même si Macron a commencé à tweeter moins fréquemment, ses tweets s'allongeaient.

Pour mieux comprendre la dynamique, j'ai répété les mêmes calculs avec 'changer' et 'continuer'. On peut trouver les chiffres et calculs en annexe 1.

Les calculs confirment que la fréquence de *devons* par tweet et par jetons de mot a diminué, mais la fréquence de *doit* par tweet a augmenté et la fréquence par jetons de mot a diminué. Cela signifie qu'il l'a utilisé davantage, mais parce que les tweets se sont allongés, la fréquence de *doit* par rapport au reste du texte a diminué.

On peut conclure que les calculs ont prouvé que depuis le mouvement, Macron a commencé à tweeter moins, puis l'utilisation de *devons* a également diminué et l'utilisation de *doit* a diminué par jetons de mot. Mais au cours de la deuxième période, quand il a tweeté, il a écrit plus longuement et on trouve plus de tweets contenant *doit*.

À mon avis, la stratégie de Macron a changé depuis le mouvement. Quand je regarde *devons*, Macron ressemble plus à un président d'action mais pas avec un fouet, plutôt comme un visionnaire parce que dans ses tweets il souligne où nous devrions être. Il sait ce que nous devons faire ou doit être mais j'ai le sentiment que le sens caché est sa vision d'un monde meilleur en quoi il croit passionnément. Et depuis le mouvement, ses croyances ont commencé à remonter encore plus parce qu'il a tweeté moins sur les petits événements et informations et encore plus sur ce que nous devrions faire. Même si les chiffres montrent qu'il a utilisé *devons* moins fréquemment, il a utilisé *doit* plus souvent, encore plus que 'doit être'. Macron a commencé à utiliser un style plus impersonnel pour prendre ses distances et ne pas sembler si agressif. Il a commencé à parler moins des choses non significatives, faisait moins de « bruit » quand il n'a pas tweeté 5 fois par jour ce qui cachait les tweets les plus significatifs, et a commencé à mettre plus d'efforts dans chaque tweet. Il adapte plutôt son langage en fonction du problème mais lui-même ne change pas.

Ce que j'ai aussi remarqué en lisant les concordances, c'est que depuis le mouvement, Macron a commencé à devenir plus philosophique. Il décrit et explique davantage, ses phrases comportent souvent une liste. On peut regarder l'exemple 4.

Ex. 4. L'Europe est un espace de paix, de prospérité, de liberté. L'Europe est le plus grand espace libre et démocratique du monde. Elle doit être fière de son modèle et le protéger. L'important est de parler de l'avenir : relation avec la Chine, climat, stratégie industrielle. (Macron 22/03/2019)

Par rapport au début du corpus, Macron a tweeté moins fréquemment mais les tweets sont devenus plus longs et plus complexes.

J'ai mentionné le sentiment qu'il laisse et je voulais continuer à l'étudier. Qui est Emmanuel Macron sur Twitter ?

3.3. Le « je » de Twitter

Président dans la vraie vie, président lors d'un discours, président sur les réseaux sociaux – le président est le même mais le discours est différent. Qui est Macron sur les réseaux sociaux ? Quel est son ethos discursif ?

Selon Nathalie Schuck (2018) Macron « est coutumier de ces envolées lyriques et tirades philosophiques stratosphériques, pour ne pas dire élitistes. » Il aime jongler avec les mots, il utilise des phrases de la littérature ce que Mayaffre appelle « l'élitisme lexical » et Longhi l'appelle le « super patron du discours » (Schuck 2018).

3.3.1. Usage le plus classique du dispositif

Le but original de Twitter selon Longhi (2013 : 1) était de partager ce qui se passe, le slogan était « Quoi de neuf ? » ou « Que se passe-t-il ? ». Pour Macron aussi, Twitter est souvent, spécialement avant le mouvement, en effet surtout un support pour ses actions. Pour décrire son style plus en détail, on peut voir la figure 7, un exemple du 23 septembre 2019.



Figure 7. Tweet d'Emmanuel Macron le 23 septembre 2019

Même si Macron est un locuteur émotif, il joue avec des mots, il tweete sur quoi faire, sur ses idées et ses croyances, son compte Twitter est encore fréquemment lié à ses discours et à son agenda. Souvent il partage en direct une vidéo de lui parlant lors d'une conférence ou faisant un discours. Il écrit un commentaire puis ajoute une URL de lui et quelquefois un hashtag.

Dans cet exemple il a dit pourquoi lui ou d'autres sont réunis là-bas – pour discuter de la protection des forêts tropicales qui absorbent le dioxyde de carbone et pour cette raison sont vitales contre le réchauffement climatique. Avec 15 mots, il a dit pourquoi ils sont là, ce qu'ils font et pourquoi c'est important. L'hashtag transportait du contenu pour comprendre où est le président et avec qui. Un tweet peut répondre à toutes les questions sur ce qu'il fait, mais ne demande pas beaucoup d'efforts.

Quand on parle des spécificités linguistiques de Macron sur Twitter, je n'ai pas trouvé des transgressions grammaticales pour raccourcir le texte mais les transgressions ne sont pas non plus recommandées pour le président. Mais il utilise de nombreux

hashtags, URLs et références à d'autres utilisateurs ou événements – technologie discursive (Paveau 2012) de Twitter, surtout quand il parle de climat.

3.3.2. « Je veux » qui se définit

Mayaffre (Lorriaux 2017) a trouvé que dans les discours Macron utilise beaucoup de *je*. Mais le discours présidentiel est un genre très différent des tweets. Les discours sont censés être longs et faire réfléchir, les tweets sont très courts et destinés à interagir rapidement, à informer. Mais selon AntCont, Macron a dit *je* 422 fois (308 avant et 114 depuis le mouvement) ce qui est intéressant car il n'y a pas besoin de dire *je* sur Twitter. Ce support sert à soutenir ses actions, à partager des informations sur les lieux, les événements, des photos, etc. ce qu'il fait aussi fréquemment. On peut regarder l'exemple 5.

Ex. 5. Je vous donne rendez-vous ce soir : je répondrai en direct aux questions de Jean-Jacques Bourdin et d'Edwy Plenel. <https://t.co/o9lJZUz8a4>. (Macron 15/04/2018)

Il ne partage souvent qu'une photo de lui, par exemple en serrant la main d'un autre homme politique avec une courte phrase comme ici dans un endroit avec quelqu'un. Mais j'ai découvert que sur Twitter, il utilise également *je* pour parler de ses idées et de sujets plus complexes.

Macron est avant tout un idéaliste qui regarde toujours en avant. Mais il n'est pas seulement dans les nuages, le *je* de Macron est confiant comme Mayaffre a dit (Lorriaux 2017), le même style est visible partout. On peut regarder l'exemple 6.

Ex. 6. Je souhaite que les femmes, partout en Afrique puissent choisir. Choisir de se marier, choisir le nombre d'enfants. Je me bats pour l'émancipation partout dans le monde et je m'engage aux côtés de Présidents qui ont fait ce choix comme le Président Issoufou. <https://t.co/YsFOvanZxV>. (Macron 23/12/2017)

Il exprime son souhait – l'émancipation des femmes africaines, ensuite affirme avec un *je* présidentiel que « je le ferai, je vais lutter contre l'injustice » et à la fin il ajoute une vidéo de lui expliquant plus précisément, généralement un extrait de son discours. C'est son style habituel.

J'ai voulu en savoir plus sur la manière dont Macron l'utilise sur Twitter, donc j'ai cherché des segments commençant par *je*.

L'un des segments les plus fréquents commençant par *je* était *je veux*. Macron l'a dit 27 fois avant le mouvement et 14 fois depuis le mouvement. Ce segment a été si peu utilisé qu'il n'y a pas de raison de le distribuer en 'changer' et 'continuer' comme avec les mots de test précédents, les résultats seraient si petits qu'ils pourraient être une coïncidence. Mais le contexte des tweets avec *je veux* après et depuis le mouvement est ce sur quoi il faut concentrer l'attention. Les tweets de Macron en décembre 2018, le pic des Gilets jaunes, sont particulièrement frappants (cf. ex. 7-9). Il est également à noter qu'avant décembre il a tweeté *je veux* en septembre et après décembre il y a eu un écart jusqu'en mars où il n'a pas du tout tweeté *je veux* alors que généralement chaque mois il produisait un ou deux tweets le contenant.

- Ex. 7. Je sais qu'il m'est arrivé de blesser certains d'entre vous par mes propos. Je veux ce soir être très clair avec vous : si je me suis battu pour bousculer le système en place, c'est justement parce que je veux servir notre pays et que je l'aime. (Macron 10/12/2018)
- Ex. 8. C'est d'abord l'état d'urgence économique et social que je veux décréter aujourd'hui pour qu'on puisse vivre mieux de son travail. (Macron 10/12/2018)
- Ex. 9. Je veux avoir un mot pour tous ceux qui au quotidien permettent à notre République d'œuvrer à la plus grande dignité de chacun. Nos militaires, nos pompiers, nos gendarmes, nos policiers, nos personnels soignants, les élus de la République, les engagés bénévoles des associations. (Macron 31/12/2018)

On peut voir que les tweets sont assez émotifs, spécialement le premier où il a l'air de s'excuser, où il dit qu'il sait avoir blessé quelqu'un ; mais il s'arrête presque une seconde, l'humilité s'estompe et on peut presque sentir sa voix retomber et il commence une nouvelle phrase avec un tout autre côté de lui – le *je* présidentiel est presque en colère, il va être très clair : il est convaincu qu'il fait la bonne chose, il ne s'excuse pas d'avoir servi son pays comme il estime est juste. Le *je veux* est certain : ce que je veux et je le ferai.

Le deuxième tweet n'est pas si émotif mais exprime à sa certitude ce qui met l'accent sur ce qu'il fera. « C'est d'abord » signifie que ce n'est pas la première de ses actions qu'il veut faire ou il a déclaré que c'était avant tout un conflit dangereux ce qui doit être traité et il le fera.

Le troisième est un mélange de gratitude, de colère et de stratégie. Il veut remercier les gens qui sauvent la dignité de la France alors que d'autres – les Gilets jaunes – la ruinent. Mais en fait il avait désespérément besoin de leur aide pour lutter contre les

Gilets jaunes parce que les masses étaient dans les rues, il ne pouvait plus faire d'erreurs. Les remerciements du président sont en fait plutôt un appel à l'aide.

Ce qui ressort de ces tweets, c'est sa certitude et sa modalité d'action – je le ferai, il croit qu'il fait la bonne chose, mais il s'inquiète comment aller en avant avec ses objectifs. Il explique plus pour demander du soutien pour survivre et pèse chaque mot pour ne pas verser d'huile sur le feu.

Quand j'ai comparé ce segment avant et depuis le mouvement des Gilets jaunes, les changements sont visibles. Avant il était très sûr de lui : je veux que la France soit, je veux que le monde soit, je veux faire, etc. Le « je veux » était plus 'doit être'. On peut regarder l'exemple 10.

Ex. 10. Je veux que la France soit mieux identifiée pour ce qu'elle est : une nation d'innovation (Macron 09/01/2018).

Il parle librement de ses souhaits et espère pour un avenir meilleur et il est certain que c'est faisable. Il est comme le motivateur des compétitions sportives qui encourage « vous pouvez le faire » faisant impression que c'est la France qu'il encourage à continuer mais en fait il n'est jamais sur le côté, il veut être à l'avant portant le flambeau.

Mais après le mouvement, quand il a commencé à tweeter moins et à changer de style, il a aussi changé sa façon de dire *je veux*. Le *je veux* s'est transformé en 'je souhaite', il est plus doux. On peut regarder l'exemple 11.

Ex. 11. Je veux que le débat que nous avons eu ce soir à Rodez continue. Une plateforme citoyenne vient d'être mise en place. Participez-y : <https://t.co/tdMhZYrvsK> <https://t.co/uCz98ZkoLO>. (Macron 03/10/2019)

Il présente une plateforme où un débat pourrait se poursuivre et souhaite à tout le monde de participer. C'est plus une invitation comme venez, s'il vous plait. En ce qui concerne *je veux*, le fort visionnaire est devenu plus modeste et humble, c'est un moyen de demander la coopération.

Conclusion

Dans ce mémoire j'ai discuté dans une perspective socio-discursive 2306 tweets d'Emmanuel Macron du 12 décembre 2017 au 11 décembre 2019 à l'aide d'outils des humanités numériques tels que Lexico 5, Voyant Tools et AntConc, afin de caractériser, sur la base des résultats de ces analyses lexicométriques, ce qu'on pourrait appeler l'ethos discursif de Macron sur Twitter.

L'un des objectifs de mon mémoire était d'introduire les humanités numériques au département pour que les futurs étudiants qui travaillent avec l'analyse du discours ou de grands corpus de texte puissent utiliser des méthodes similaires et travailler plus rapidement. Les humanités numériques sont conçues pour fonctionner sur de grands corpus de texte, qu'il est inutile d'analyser à la main. Ils aident à trouver des modèles que le cerveau humain traite plus lentement. Avec l'aide de mon travail, les étudiants peuvent suivre ma trajectoire et simplifier leur. J'ai présenté trois logiciels, j'ai décrit comment je les ai utilisés et je les ai évalués en tant que débutante dans ce domaine.

Ce qui est fascinant pour moi avec les logiciels en humanités numériques, c'est qu'ils peuvent presque prédire l'avenir (au sens de calculer les probabilités), si la bonne quantité de données est insérée. Il me semble qu'il est ainsi possible d'analyser une thématique ou un locuteur lors de la recherche de motifs, et de conclure puis comment il pourrait agir dans une situation similaire.

Dans le cadrage théorique, j'ai présenté ce que sont les humanités numériques, comment ce type de méthodologie a déjà été utilisé dans l'analyse française du discours où diverses méthodes basées sur les machines pour mesurer l'impact social des mots dans le discours ont été utilisées et comment la lexicométrie a été utilisée pour étudier les discours présidentiels. J'ai discuté du travail de Damon Mayaffre (2004 ; 2007 ; 2017 selon Lorriaux ; 2020), qui a analysé les récurrences comparatives et les motifs de vocabulaire dans les discours publics de tous les présidents et du travail de Julien Longhi (2013 ; *et al* 2017 ; 2017), qui a étudié le fonctionnement des discours sur Twitter et a lancé une plate-forme #Idéo2017 pour analyser à l'aide d'outils de la lexicométrie ce que les candidats présidentiels disent sur les réseaux sociaux. Tous deux ont étudié Emmanuel Macron, mais différemment que mon mémoire car Mayaffre a étudié les discours publics et Longhi a étudié Macron avant les élections en le comparant avec d'autres candidats. Dans la deuxième partie du premier chapitre,

j'ai présenté Twitter où mon corpus est collecté. J'ai caractérisé Twitter comme une plateforme et les tweets politiques comme objet d'étude.

Le deuxième chapitre s'est concentré sur mon corpus et sur le premier test avec les outils lexicométriques. J'ai expliqué comment j'ai collecté et filtré les tweets, puis j'ai présenté les programmes et les premiers résultats.

Les dix noms et adjectifs les plus utilisés par Macron par ordre décroissant étaient *France, pays, français, monde, ensemble, ans, aujourd'hui, république, européenne* et *nouvelle*. Les verbes les plus utilisés étaient *avoir, être, faire* et *devoir* – dans sa première forme plurielle « devons », ce que selon Damon Mayaffre (2004 : 38-39), aucun président depuis de Gaulle à Chirac n'a utilisé aussi fréquemment. C'est un fait lexical que j'ai donc choisi comme point de départ pour mon mémoire.

J'ai confirmé mon hypothèse d'après laquelle il est possible de décrire une personne à l'aide de programmes d'humanités numériques sans la comparer à quelqu'un d'autre, mais pour pouvoir généraliser le comportement d'une personne, il faut comparer le sujet avec lui-même. J'ai divisé mon corpus en deux pour tester comment Macron a changé son style sur Twitter depuis le mouvement des Gilets jaunes, qui se sont battus contre lui et sa politique.

Pour interpréter les tweets de manière plus approfondie, afin de caractériser l'ethos discursif de Macron, j'ai analysé le rapport que Macron avait à sa manière d'énoncer. Avec l'aide de la lexicométrie, j'ai analysé plus spécifiquement les mots clés relevés dans les sondages : *devons* et puis *je veux* sur Twitter. Pour élucider ce qu'entend Macron par *devons*, j'ai utilisé un axe sémantique spécifique, où j'ai distribué les tweets contenant le mot recherché en deux catégories : changer et continuer. Pour prendre davantage en compte les variations que permet Twitter comme le dispositif d'énonciation, j'ai calculé la longueur des tweets et je les ai décrits aussi qualitativement.

J'ai trouvé qu'avant le mouvement des Gilets jaunes, Macron a tweeté 5,15 fois par jour et depuis le début du mouvement, seulement 1,423 fois par jour – la fréquence des tweets a diminué de 3,619 fois. Mais quand on regarde la longueur des tweets, mes calculs ont confirmé que même si Macron a commencé à tweeter moins fréquemment, ses tweets s'allongeaient : avant, la longueur moyenne du tweet était de 28,571 jetons

des mots et depuis le mouvement, la longueur est passée à 42,283 jetons des mots en moyenne.

En me concentrant littéralement sur la construction de son “je” sur Twitter, j’ai constaté que Twitter comme mode de communication pour Emmanuel Macron est souvent un support pour ses actions : il diffuse des informations sur le lieu où il se trouve, sur ce qu’il fait, il ajoute une photo ou une vidéo d’une conférence et un bref commentaire. Mais en même temps, il parle librement de ses souhaits, ses tweets sont souvent philosophiques et émotifs, il joue avec les mots, il garde ses distances. Son éthos discursif montre qu’il sait ce qu’il fait et qu’il le fait avec son cœur, qu’il est passionné et déterminé et pas dans les nuages. Il n’est pas seulement homme des paroles, il promet : « je le ferai ». Il parle d’un avenir meilleur et comment *nous* pouvons y parvenir ensemble, euro-optimiste comme il est, mais il veut vraiment être celui qui porte le flambeau.

Quand un problème est survenu, dans mon corpus il s’agit du mouvement des Gilets jaunes, Emmanuel Macron a adapté son style. Il a donné l’impression de rester plus calme et a commencé à dire moins « devons » et « doit ». Il a tweeté moins pour ne pas attirer l’attention sur ce qui compte vraiment et a allongé les tweets importants. Sur la base de mes résultats, le style de Macron a un peu changé, mais lui-même est resté le même.

Mes résultats ne donnent certes pas une vue d’ensemble sur Emmanuel Macron, on n’obtient que les premières pistes. Pour la recherche académique, je suggérerais un corpus au moins trois fois plus grand parce que plus le corpus est grand, plus les résultats sont statistiquement significatifs et plus il y a à découvrir et tirer des conclusions.

La question à laquelle il reste à répondre est : pourquoi la lexicométrie ? Qu’est-ce que cette approche m’a donné ? Je comparerais les humanités numériques avec un jeu de bataille navale. Dans le jeu original, il est facile de trouver un navire parce que l’aire de jeu est limitée ; il suffit de faire un seul tir droit pour être sûr qu’il y a un bateau. Dans la vraie vie, il n’y a pas de limites, on peut tirer autant qu’on veut, mais on ne découvre jamais rien. Il faut insérer un bon mot-clé pour trouver un modèle car tous les mots ne mènent pas à une direction espérée. J’ai testé différents mots, certains ont

conduit à une conclusion, d'autres n'étaient pas assez fréquents pour faire une supposition ou ne menaient nulle part, alors je les ai supprimés.

Mais il est plus facile de tirer un grand nombre de balles à la fois avec un logiciel capable que manuellement une par une. Pour trouver des mots spécifiques du corpus, il est facile d'utiliser la commande Ctrl+F de l'ordinateur, mais quand on ne sait pas quoi chercher, les outils numériques viennent à la rescousse car ils aident à trouver des modèles qui peuvent passer inaperçus à l'œil nu. Cela aurait été inutilement long de commencer à compter tous les mots pour découvrir les plus fréquents. Les humanités numériques donnent la base – aident à voir les modèles, les changements, les fréquences, les relations, etc. – sur laquelle aller de l'avant.

Références

1. #Idéo2017 plateforme. 2017. En ligne <http://ideo2017.ensea.fr/plateforme/>. Consulté le 27 avril 2020.
2. About Twitter. 2020. En ligne <https://about.twitter.com/>. Consulté le 23 février 2020.
3. BÉNIS, O. ; GRAZIANI, C. 2020. « Coronavirus : comment Emmanuel Macron a changé sa façon de tweeter », in *France inter*. En ligne https://www.franceinter.fr/politique/coronavirus-comment-emmanuel-macron-a-change-sa-facon-de-tweeter?fbclid=IwAR0uaAsBS0LtnQ02nCT-PA2cUhr0RVh2vLKLfEGdXn2qrk_qvsiei2wU34o. Consulté le 30 mars 2020.
4. BOUZIDI, L. ; BOULESNANE, S. 2017. « Les humanités numériques : L'évolution des usages et des pratiques », in *Les Cahiers du numérique*, 13(3), p. 19-38. DOI : 10.3166/lcn.13.3-4.19.38
5. CHHABRA, R. 2020. « Twitter Diplomacy : A Brief Analysis », in *ORF Issue Brief*, 335. En ligne <https://www.orfonline.org/research/twitter-diplomacy-a-brief-analysis-60462/?fbclid=IwAR1j5OH3NhVUIDptGLUr6PgmctMl51PCJWzRHYNvEZvuevbeFOWHusE489QM>. Consulté le 14 mars 2020.
6. DACOS, M. ; MOUNIER, P. 2014. *Humanités numériques : État des lieux et positionnement de la recherche française dans le contexte international*. En ligne <https://hal.archives-ouvertes.fr/hal-01228945/document>. Consulté le 26 mars 2020.
7. DEVELY, A. 2018. Interview avec Damon Mayaffre « Emmanuel Macron ne converse pas avec le peuple, il le met à distance », in *Le Figaro*, le 11 décembre. En ligne <https://www.lefigaro.fr/langue-francaise/actu-des-mots/2018/12/11/37002-20181211ARTFIG00142-emmanuel-macron-ne-converse-pas-avec-le-peuple-il-le-met-a-distance.php>. Consulté le 20 mars 2020.
8. DODUIK, N. 2019. *Les humanités numériques, une révolution ?*. En ligne <https://hal.archives-ouvertes.fr/hal-01948447/document>. Consulté le 26 mars 2020.
9. DOURY, M. 2016. *Argumentation : analyser textes et discours*. Paris : Armand Colin.

10. Encyclopædia Britannica. 2020. *Twitter. Microblogging service*. En ligne <https://www.britannica.com/topic/Twitter>. Consulté le 22 février 2020.
11. ENGELSEN, I. A. 2014. *Le verbe modal devoir et la modalité déontique dans le discours portant sur le changement climatique*. Mémoire de master. Université de Bergen. En ligne <http://bora.uib.no/handle/1956/7977>. Consulté le 20 avril 2020.
12. GABOULAUD, A. ; LECHEVALLIER. A.-S. 2017. « Exclusif : sept mois de présidence Macron mot à mot », in *Paris Match*, le 30 décembre. En ligne <https://www.parismatch.com/Actu/Politique/Exclusif-sept-mois-de-presidence-Macron-mot-a-mot-1428923>
13. GAUMONT, N. ; PANAHI, M. ; CHAVALARIAS, D. 2018. *Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election*. DOI : 10.1371/journal.pone.0201879
14. Get Tweet Timelines. 2020. *Developer Twitter: Get Tweet Timelines*. En ligne <https://developer.twitter.com/en/docs/tweets/timelines/overview>. Consulté le 22 février 2020.
15. KÄSPER, M. 2017. *Présenter des ouvrages académiques en sciences humaines et sociales en français et en estonien. Pistes contrastives pour étudier les cultures discursives dans les comptes rendus de lecture français et estoniens*. En ligne <http://hdl.handle.net/10062/57816>. Consulté le 27 mars 2020.
16. KRONNING, H. 1996. *Modalité, cognition et polysémie : sémantique du verbe modal DEVOIR*. Uppsala : Acta Universitatis Upsaliensis.
17. LAURENCE, A. 2019. *AntConc (3.5.8). Le logiciel*. En ligne <https://www.laurenceanthony.net/software>. Consulté le 13 mars 2020.
18. LE QUERLER, N. 1996. *Typologie des modalités*. Caen : Presses Universitaires de Caen.
19. LEBLANC, J.-M. 2016. *Analyses lexicométriques des vœux présidentiels*. London : ISTE Group.
20. LEHUT, T. 2019. « Un an de Gilets jaunes : les dix dates-clés du mouvement », in *France Bleu*, le 16 novembre. En ligne <https://www.francebleu.fr/infos/societe/infographie-un-de-gilets-jaunes-les-dix-dates-cles-du-mouvement-1573744994>. Consulté le 1 avril 2020.
21. Lexico 5. 2019. En ligne <http://www.lexi-co.com/index.html>. Consulté le 13 mars 2020.

22. LONGHI, J. ; MARINICA, C. ; HASSINE, N. ; ALKHOULI, A. ; BORZIC, B. 2017. « The #Ideo2017 Platform », in *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17)*, p. 46-51. DOI : 10.5281/zenodo.1040875
23. LONGHI, J. 2013. « Essai de caractérisation du tweet politique », in *L'information grammaticale*, 136, p. 25-32. Halshs-00940202.
24. LONGHI, J. 2017. « Humanités, numérique : des corpus au sens, du sens aux corpus », in *Questions de communication*, 31, p. 7-17. En ligne <http://journals.openedition.org/questionsdecommunication/11039>. Consulté le 4 mai 2020.
25. LONGHI, J. 2017. « Le discours d'Emmanuel Macron, construction d'un storytelling », in *The Conversation*, le 26 janvier. En ligne <http://theconversation.com/le-discours-demmanuel-macron-construction-dun-storytelling-71847>. Consulté le 26 mars 2020.
26. LORRIAUX, A. 2017. « Le « je » d'Emmanuel Macron », in *Sciences Humaines*, 295, l'août-septembre. En ligne https://www.scienceshumaines.com/le-je-d-emmanuel-macron_fr_38469.html. Consulté le 20 mars 2020.
27. MAINGUENEAU, D. 2016. *Énonciation et analyse du discours*. Corela HS-19, DOI : 10.4000/corela.4446
28. MALDIDIER, D. 1969. *Analyse linguistique du vocabulaire politique de la guerre d'Algérie d'après six quotidiens parisiens*. En ligne http://classiques.uqac.ca/contemporains/malidier_denise/analyse_linguistique/analyse_linguistique.html. Consulté le 26 mars 2020.
29. Manuel AntConc version 3.5.8. 2019. En ligne <https://www.laurenceanthony.net/software/antconc/releases/AntConc358/help.pdf>. Consulté le 13 mars 2020.
30. MARCELLESI, J.-B. 1971. « Éléments pour une analyse contrastive du discours politique », in *Langages*, 23, « Le discours politique », p. 25-56. En ligne https://www.persee.fr/doc/lgge_0458-726x_1971_num_6_23_2049. Consulté le 20 mars 2020.
31. MAYAFFRE, D. 2004. *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la Vème République*. Paris : Champion.
32. MAYAFFRE, D. 2007. « L'analyse de données textuelles aujourd'hui : du corpus comme une urne au corpus comme un plan : Retour sur les travaux actuels de

- topographie/topologie textuelle », in *Lexicometrica*, p. 1-12. En ligne <https://hal.archives-ouvertes.fr/hal-00551468>. Consulté le 20 mars 2020.
33. MAYAFFRE, D. ; GAURESI, M. ; VANNI, L. 2020. « Ces mots que Macron emprunte à Sarkozy. Discours et intelligence artificielle », in *Dispositifs numériques et dévoilement de soi*, 21, le 9 mars. En ligne <http://journals.openedition.org/corpus/5105>. Consulté le 20 mars 2020.
 34. MEUNIER, J.-G. 2017. « Humanités numériques et modélisation scientifique », in *Questions de communication*, 31, p 19-48. DOI : 10.4000/questionsdecommunication.11040
 35. MONGAILLARD, V. 2018. « Tics de langage : « En même temps », le péché mignon de Macron », in *Le Parisien*, le 21 avril. En ligne <http://www.leparisien.fr/societe/tics-de-langage-en-meme-temps-le-peche-mignon-de-macron-21-04-2018-7676160.php>
 36. NARUSK, K. 2014. *Les équivalents estoniens du verbe modal devoir*. En ligne <http://hdl.handle.net/10062/41811>. Consulté le 20 avril 2020.
 37. PATIN, S. ; MEGALE, F. 2018. « Observing Eurolects : The case of French » in *Corpus analysis of linguistic variation in EU law*, p. 122-146. DOI : 10.1075/sc1.86
 38. PAVEAU, M.-A. 2012. « Activités langagières et technologie discursive. L'exemple de Twitter », in *La pensée du discours*, le 27 février. En ligne <http://penseedudiscours.hypotheses.org/8338>. Consulté le 23 février 2020.
 39. PÊCHEUX, M. 1969. *Analyse automatique du discours*. Paris : Dunod.
 40. Perspectives SSF. 2014. *Sciences humaines numériques (digital humanities)*. En ligne <https://www.usherbrooke.ca/ssf/veille/perspectives-ssf/numeros-precedents/octobre-2014/le-fin-mot-sciences-humaines-numeriques/>. Consulté le 18 mars 2020.
 41. R Project. 2020. *The R Project for Statistical Computing. Getting Started*. En ligne <https://www.r-project.org/>. Consulté le 22 mars 2020.
 42. SALEM, A. 1987. *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
 43. SCHUCK, N. 2018. « Parlez-vous le Macron ? », in *Le Parisien*, le 2 février. En ligne <http://www.leparisien.fr/politique/do-you-speak-macron-02-02-2018-7538088.php>. Consulté le 13 mars 2020.
 44. SINCLAIR, S. ; ROCKWELL, G. 2016. *Voyant Tools*. En ligne <http://voyant-tools.org/>. Consulté le 13 mars 2020.

45. TULIS, J. K. 1987. *The rhetorical presidency*. Princeton : Princeton University Press.
46. VETTERS, C. 2012. « Modalité et évidentialité dans pouvoir et devoir : typologie et discussions », in *Langue Française*, 173, p. 31-47. DOI : 10.3917/lf.173.0031
47. Voyant Tools Help. Guides. 2016. En ligne <https://voyant-tools.org/docs/#!/guide/start>. Consulté le 13 mars 2020.
48. WANG, C. 2019. « Twitter Diplomacy : Preventing Twitter Wars from Escalating into Real Wars », in *Future of Diplomacy Project*, le 20 mai. En ligne <https://www.belfercenter.org/publication/twitter-diplomacy-preventing-twitter-wars-escalating-real-wars?fbclid=IwAR3IeOk0LEWxgL6o8VHUMruQLmrnBvhYPpsi1eGDhaIyhyjmEcsPIBW3LoE>. Consulté le 14 mars 2020.

Resümee

„Leksikomeetria Emmanuel Macroni säutsude analüüsimiseks“

Käesolev bakalaureusetöö uurib, kuidas kasutab Prantsusmaa president Emmanuel Macron Twitterit ja millise kuvandi ta endast sotsiaalmeedias loob. Töö eesmärgiks on analüüsida presidendi sõnakasutust, kasutades tekstikorpusi võrdlevaid leksikomeetrilisi programme. Autor väidab, et digihumanitaaria abil on võimalik kirjeldada kõnelejat ilma kellegi teisega võrdlemata, leida keelest iseloomulikke kordusi ja kirjeldada kõnelejale omaseid diskursiivseid käitumismustreid, et nende põhjal ennustada, kuidas käitub kõneleja sarnases situatsioonis.

Kuna digihumanitaaria on romanistika osakonnas uudne, kirjeldab autor digirakenduste kasutamist kui oma teekonda, mis oleks kergesti järgitav ning abivahendiks ka teistele õpilastele. Siinkohal on oluline mainida, et autori enda jaoks oli digivahendite kasutamine esmakordne, mistõttu on tegemist eelkõige esimeste katsetustega. Küll aga juhib autor tähelepanu võimalustele, mida digihumanitaaria pakub.

Töö on jaotatud kolmeks peatükiks. Esimeses peatükis tutvustatakse digihumanitaaria ja prantsuse diskursusanalüüsi teoreetilist raamistikku ning sotsiaalmeediaplatformi Twitter, keskendudes sellele, kuidas erineb poliitiline säuts avalikust kõnest. Teine peatükk kirjeldab uuritava tekstikorpuse moodustamist ja digivahendite rakendamist. Autor annab programmidele oma kogemuse põhjal hinnangu ja arutleb esimeste tulemuste üle. Teine peatükk on mõeldud järgnevatele põlvvedele kasutamiseks, et teha edaspidi teadlikumaid valikuid. Kolmas peatükk läheb saadud andmetega süvitsi ja keskendub nende semantilisele tõlgendamisele, iseloomustamaks president Macroni kuvandit Twitteris. Töö analüüsiosa on üles ehitatud vaid esimestele tulemustele, kuid võimalusi on lugematu hulk.

Tekstikorpuse töötlemiseks katsetati kolme digirakendust: 1) Lexico 5, mis on prantsuse koolkonnas leksikomeetria uurimiseks sümboolse tähendusega; 2) Voyant Tools, mis on veebipõhiselt kättesaadav ja loob automaatselt mitmesuguseid visuaale, mis aitavad tekstikorpust mõtestada ning 3) AntConc, mis on professionaalseks kasutamiseks, kuid algajale kergesti mõistetav.

Bakalaureusetöös uuritav korpus koosnes 2306st eeldatavasti Emmanuel Macroni kirjutatud säutsust ajaperioodil 12. detsember 2017 kuni 11. detsember 2019. Selleks, et kirjeldada presidendi muutumist ajas, jagas autor vaadeldava perioodi kaheks: enne ja pärast Kollavestide liikumist, mis sai alguse 17. novembril 2018.

Uurimise käigus selgus, et alates Kollavestide liikumisest vähenes presidendi aktiivsus Twitteris 3,6 korda, kuid tema säutsud muutusid pikemaks – kui enne oli keskmine säuts 28,6 sõna pikk, siis alates Kollavestide liikumisest oli keskmine säuts 42,3 sõna pikk.

Semantilisemaks tõlgendamiseks täiendati modaalsusanalüüsi spetsiaalselt korpuse jaoks loodud semantilise teljega Presidendi kuvandit kirjeldati kolme lähemalt vaadeldud sõna ja segmendi abil: *devons* ('me peame'), *doit* ('peab') ja *je veux* ('mina tahan').

Bakalaureusetöö tulemusena selgus, et president Emmanuel Macroni diskursuslik eetos on eelkõige 'mina teen', kus *mina* tähistab presidenti. Macron ütleb eelnevatest presidentidest tihedamini *mina* ning sealjuures, mida ta tahab või soovib. Ta on emotsionaalne kõneleja, kelle säutsud on sageli filosoofilised, kohati isegi keerulised mõista, mis on tema teadlik valik, et end lugejast distantseerida. Alates Kollavestide liikumisest muutus tema stiil aga leebemaks – Macroni 'mina tahan' omandas semantiliselt rohkem tähenduse 'mina soovin' ning tema säutsud muutusid kirjeldavamaks, mida ja miks ta täpselt soovib. Saadud tulemuste põhjal autor väidab, et alates Kollavestide liikumisest on Emmanuel Macroni säutsumise stiil muutunud, kuid tema ise on jäänud samaks. Macron kohandab strateegiliselt oma käitumist vastavalt olukorrale.

Annexe 1. Les résultats en chiffres et les calculs

On trouve ici les tableaux de mes résultats de distribution et les calculs de fréquences, divisé avant et depuis le mouvement des Gilets jaunes. Le premier tableau contient les fréquences de *devons*, *devons changer* et *devons continuer*. Le deuxième tableau contient les fréquences de *doit*, *doit changer* et *doit continuer*.

Plage de temps	Avant le mouvement	Depuis le mouvement
Le nombre total de <i>devons</i>	104	25
Le nombre de <i>devons changer</i>	59	13
Le nombre de <i>devons continuer</i>	45	12
La fréquence de <i>devons</i> par tweet	$104/1751=0,059$	$25/555=0,045$
La fréquence de <i>devons changer</i> par tweet	$59/1751=0,034$	$13/555=0,023$
La fréquence de <i>devons continuer</i> par tweet	$45/1751=0,026$	$12/555=0,022$
La fréquence de <i>devons</i> par jetons de mot	$104/50027=0,00208$	$25/23467=0,00107$
La fréquence de <i>devons changer</i> par jetons de mot	$59/50027=0,00118$	$13/23467=0,00055$
La fréquence de <i>devons continuer</i> par jetons de mot	$45/50027=0,00090$	$12/23467=0,00051$

Plage de temps	Avant le mouvement	Depuis le mouvement
Le nombre total de <i>doit</i>	71	24
Le nombre de <i>doit changer</i>	44	15
Le nombre de <i>doit continuer</i>	27	9
La fréquence de <i>doit</i> par tweet	$71/1751=0,041$	$24/555=0,043$
La fréquence de <i>doit changer</i> par tweet	$44/1751=0,025$	$15/555=0,027$
La fréquence de <i>doit continuer</i> par tweet	$27/1751=0,015$	$9/555=0,016$
La fréquence de <i>doit</i> par jetons de mot	$71/50027=0,00142$	$24/23467=0,00102$
La fréquence de <i>doit changer</i> par jetons de mot	$44/50027=0,00088$	$15/23467=0,00064$
La fréquence de <i>doit continuer</i> par jetons de mot	$27/50027=0,00054$	$9/23467=0,00038$

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Liina Maurer,

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Lexicométrie pour analyser les tweets d’Emmanuel Macron“, mille juhendaja on Marge Käsper, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Liina Maurer

24.04.2020